



# Tetramer- Dependence of DNA Conformation

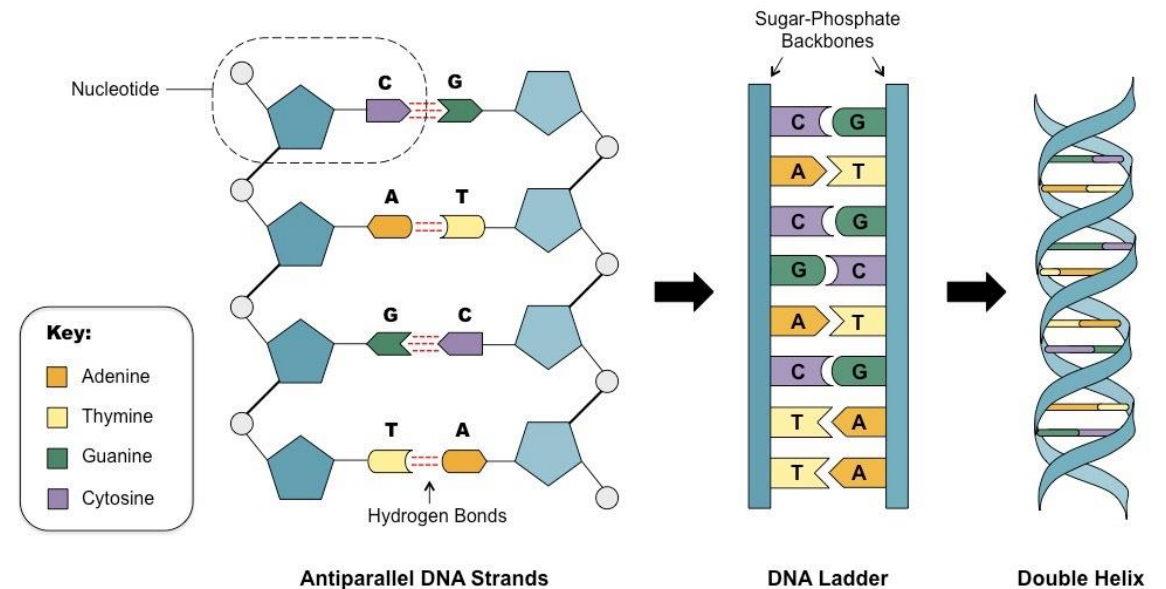
---

ZOE WEFERS

MENTOR: DR. WILMA OLSON

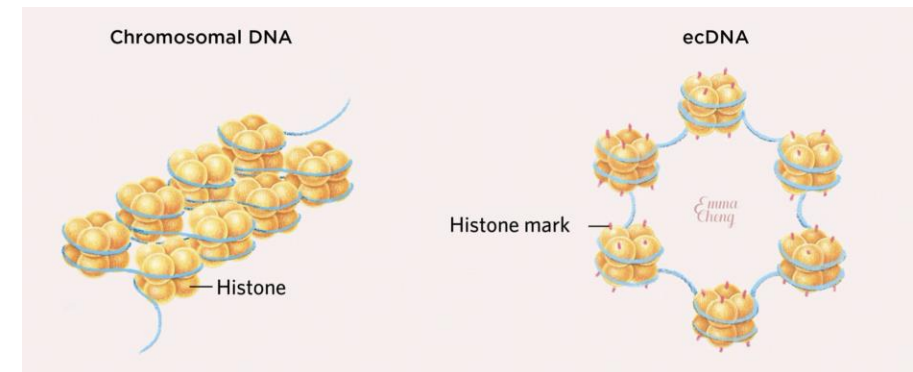
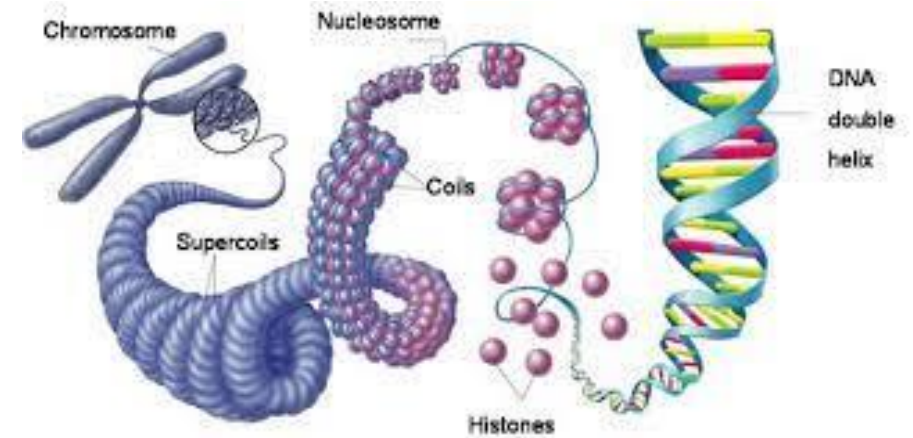
# DNA: A Biological View

- DNA is a biological molecule that encodes genetic information
- 4 types of nucleic acid bases, adenosine (A), cytosine (C), guanosine (G), thymidine (T)
- Nucleic acid bases form base pairs
  - A pairs with T and C pairs with G
- Read one strand from 5' to 3' - ex: CACGACTT
- Double helix structure - stiff
  - Bending and twisting affects elastic energy



# DNA: The Important of Folding

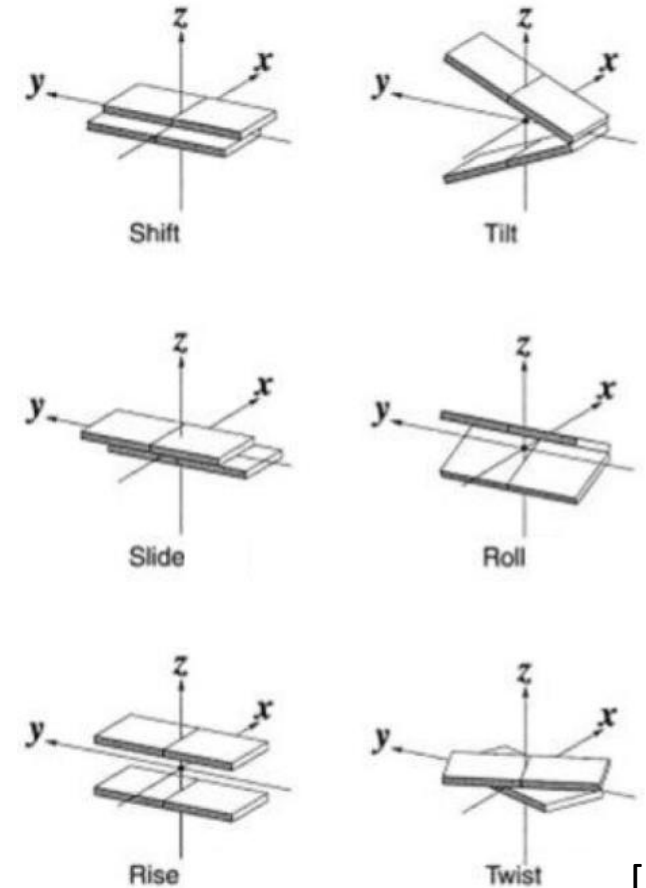
- Packaged in nucleus by wrapping around histone proteins
- DNA conformation can influence gene regulation
  - Ex: transcription
- Extrachromosomal circular DNA is found in some cancers
  - Why we study DNA minicircles



# DNA: A Mathematical View

- Model each base pair as rigid rectangle
- Orientation between adjacent base pairs described represented by 6 "base pair step parameters": Tilt, Roll, Twist, Shift, Slide, Rise
  - $\underline{p} = (\theta_1, \theta_2, \theta_3, \rho_1, \rho_2, \rho_3)$
  - Intrinsic parameter, denoted  $\underline{p}_0 =$  Rest state step parameters
- Stiffness matrix ( $\mathbf{F}$ ) = covariance matrix inverse

$$\mathbf{F}^{-1} = \begin{bmatrix} \langle \theta_1^2 \rangle - \langle \theta_1 \rangle^2 & \langle \theta_1 \theta_2 \rangle - \langle \theta_1 \rangle \langle \theta_2 \rangle & \dots & \dots & \dots & \langle \theta_1 \rho_3 \rangle - \langle \theta_1 \rangle \langle \rho_3 \rangle \\ \langle \theta_1 \theta_2 \rangle - \langle \theta_1 \rangle \langle \theta_2 \rangle & \langle \theta_2^2 \rangle - \langle \theta_2 \rangle^2 & \dots & \dots & \dots & \langle \theta_2 \rho_3 \rangle - \langle \theta_2 \rangle \langle \rho_3 \rangle \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \langle \theta_1 \rho_3 \rangle - \langle \theta_1 \rangle \langle \rho_3 \rangle & \langle \theta_2 \rho_3 \rangle - \langle \theta_2 \rangle \langle \rho_3 \rangle & \dots & \dots & \dots & \langle \rho_3^2 \rangle - \langle \rho_3 \rangle^2 \end{bmatrix}$$



# Elastic Energy

---

$$\underline{p}^i = (\theta_1^i, \theta_2^i, \theta_3^i, \rho_1^i, \rho_2^i, \rho_3^i)$$

Step Parameters for the ith step

$$E^i = \frac{1}{2} (\underline{p}^i - \underline{p}_0^i)^T \mathbf{F}^i (\underline{p}^i - \underline{p}_0^i)$$

Elastic Energy for the ith step

$$E = \sum_{i=1}^{N-1} E^i$$

Total Elastic

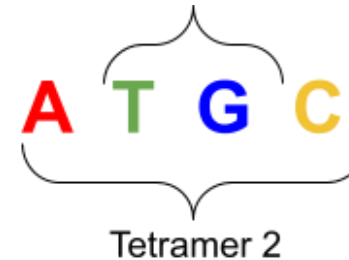
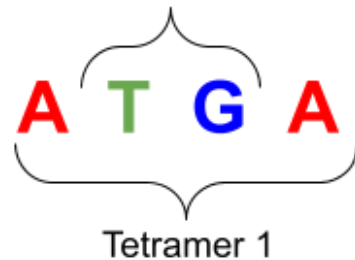
$$\frac{\partial E^i}{\partial \underline{p}^i} = \mathbf{F}_s^i (\underline{p}^i - \underline{p}_0^i)$$

Derivative of Elastic Energy

# What is a Sequence Dependent Model?

---

- Find intrinsic parameter and stiffness matrices using either high resolution structural data or molecular dynamics simulation data
- Dimer-dependent Model: Intrinsic parameters depends on the base pairs in the step
  - ie intrinsic parameters for AA step different than GC step
- Tetramer Dependent model: Intrinsic parameters depends on tetramer centered at the step
  - New evidence suggests intrinsic parameter are affected also by flanking base pairs

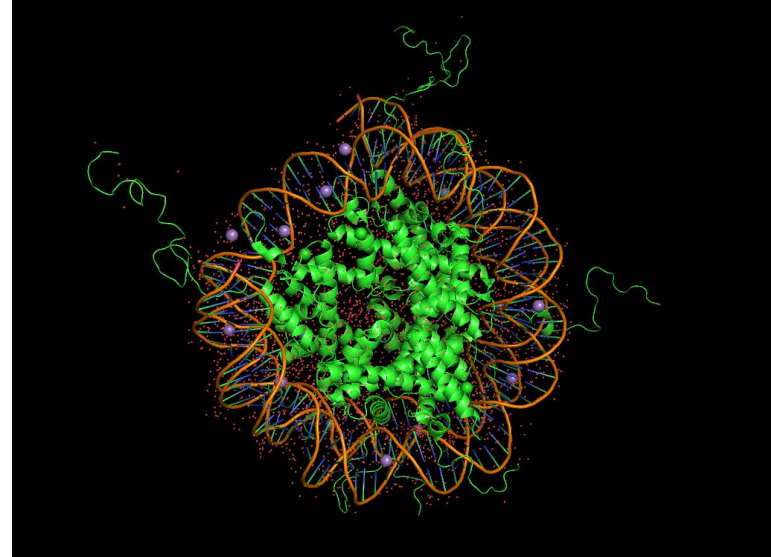
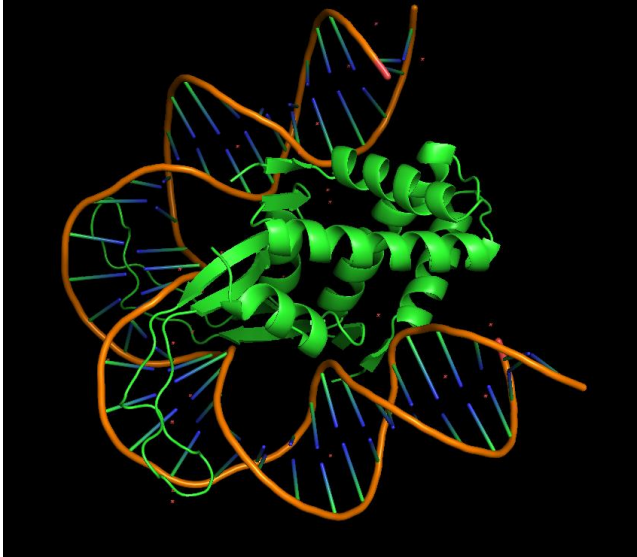


- Models we used: IdealDNA (sequence-independent), Olson1998 (dimer), Cohen2017\_dim (dimer), Cohen2017\_tet (tetramer)

# emDNA Software – Current Functionalities

---

- Core Idea: DNA will move to achieve least elastic energy
  - Can be modeled by gradient descent-like algorithm
- **User has the option to use dimer-dependent models or sequence independent model**
  - **But no tetrameric functionality**



## My Goals:

1 ) Adapt emDNA to allow user to choose to tetramer-dependent, dimer-dependent or sequence independent model to optimize elastic energy

2) Compare the intrinsic parameters of new Cohen2017 sequence dependent models

3) Use the modified program to see how using different sequence dependent models changes the minimum energy conformations achieved



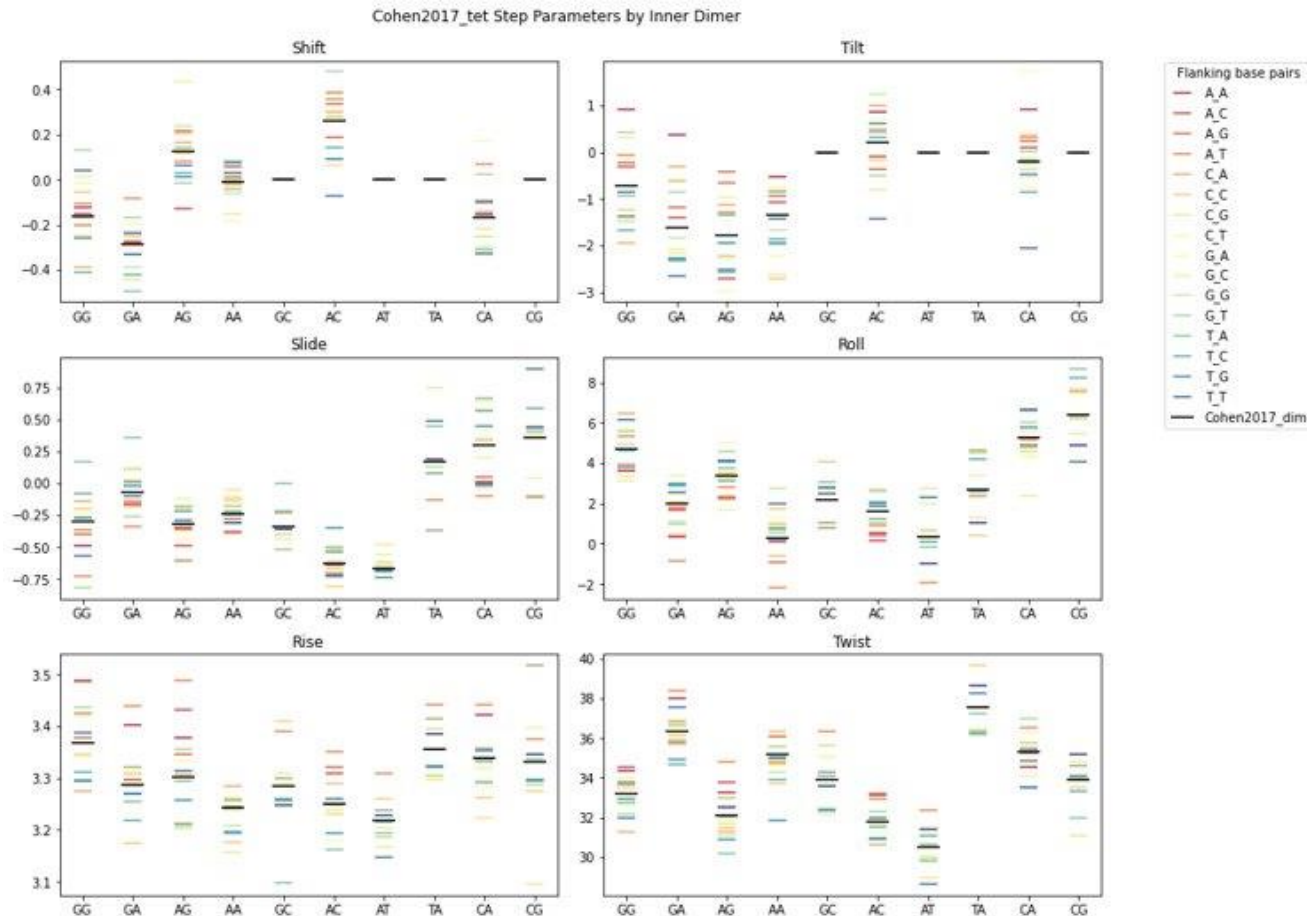
# Goal 1: Adapting emDNA

- Successfully implemented the tetrameric functionality!

```
Sequence.cpp — emDNA
C StepParameters_Cohen2017_tet.h  Sequence.cpp 9+ X  StepParametersDB.cpp  BpColl
DNASim > src > dna > Sequence.cpp > {} DNASim > TetramerSequence::third_base() const
66
67 // Added by Zoe Wefers (McGill University, June 2021, DIMACS REU)
68 /** TetramerSequence class */
69 TetramerSequence::TetramerSequence(const BaseSymbol& base_1,
70                                     const BaseSymbol& base_2,
71                                     const BaseSymbol& base_3,
72                                     const BaseSymbol& base_4) :
73     m_bases(base_1, base_2, base_3, base_4) {}
74
75 //base accessors
76 const BaseSymbol& TetramerSequence::first_base() const {
77     return std::get<0>(m_bases);
78 };
79 const BaseSymbol& TetramerSequence::second_base() const {
80     return std::get<1>(m_bases);
81 };
82 const BaseSymbol& TetramerSequence::third_base() const {
83     return std::get<2>(m_bases);
84 };
85 const BaseSymbol& TetramerSequence::fourth_base() const {
86     return std::get<3>(m_bases);
87 };
```

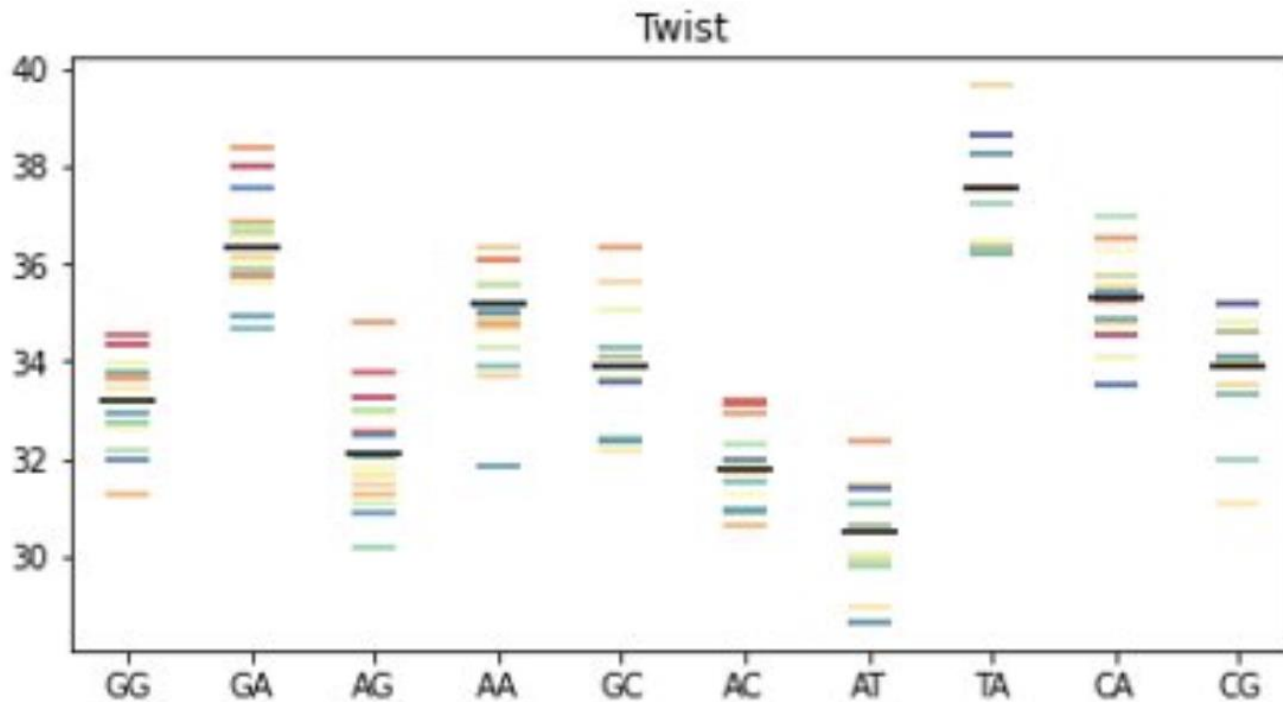
```
StepParameters_Cohen2017_tet.h — emDNA
C StepParameters_Cohen2017_tet.h  StepParametersDB.cpp  BpCollection.cpp  SequenceDepenceModels.h
DNASim > src > dna > C StepParameters_Cohen2017_tet.h > ...
1 // StepParameters_Cohen2017_tet data file
2 // Zoe Wefers (McGill University, June 2021, DIMACS REU)
3
4
5 // Cohen2017_tet intrinsic step parameters
6
7
8 #ifndef StepParameters_Cohen2017_tet_h
9 #define StepParameters_Cohen2017_tet_h
10
11
12 const std::string StepParameters_Cohen2017_tet[400] = {
13
14     "AAAA={-0.53345, -0.60993, 36.33769, -0.0103, -0.17685, 3.25888}",
15     "AAAC={-1.07744, 0.0094, 34.75171, 0.05559, -0.37426, 3.25636}",
16     "AAAG={-0.92824, -0.92492, 36.09203, 0.01365, -0.27688, 3.26021}",
17     "AAAT={-1.34306, -2.17698, 36.05945, -0.00149, -0.38719, 3.24301}",
18     "CAAA={-0.82776, 0.9711, 34.79466, -0.02238, -0.12939, 3.28412}",
19     "CAAC={-2.70954, 1.72873, 33.70738, -0.04069, -0.1098, 3.17598}",
20     "CAAG={-2.6339, 1.01149, 34.64909, -0.15033, -0.05463, 3.15741}",
21     "CAAT={-2.21904, 1.30474, 35.65932, -0.18271, -0.12259, 3.25249}",
22     "GAAA={-0.74971, -0.59721, 36.32128, 0.00168, -0.15655, 3.26335}",
23     "GAAC={-1.33428, 0.65572, 34.94217, 0.03676, -0.17502, 3.23823}",
24     "GAAG={-1.67994, 2.73603, 34.2923, -0.06342, -0.22697, 3.20840}",
25     "GAAT={-1.97544, 0.56103, 35.56169, -0.04375, -0.30444, 3.25858}",
26     "TAAA={-0.86074, 0.69529, 35.04674, 0.00803, -0.24754, 3.24441}",
27     "TAAC={-1.85495, 0.79784, 33.88869, 0.08133, -0.22325, 3.1946}",
28     "TAAG={-1.94496, 0.24801, 31.81883, 0.07445, -0.30508, 3.19623}",
29     "TAAT={-1.41621, 1.99504, 34.95682, 0.03138, -0.30843, 3.24369}",
30     "AACA={0.85246, 0.5373, 33.14662, 0.33208, -0.63701, 3.28978}",
31     "AACG={-0.09784, 0.16398, 33.20058, 0.38386, -0.71941, 3.30819}",
32     "AACG={-0.36144, 0.40901, 33.11126, 0.1874, -0.70753, 3.32139}",
33     "AACT={0.47387, 0.01002, 32.9205, 0.35835, -0.81184, 3.35028}",
34     "CACA={0.99098, 1.04248, 30.62603, 0.30199, -0.52005, 3.23026}";
```

# Goal 2: Comparing Cohen2017 models



- Significant effect of flanking base pairs provides evidence that potential differences in optimized configuration of are in fact caused by dimer vs tetramer models
  - Not by other computational factors

# Goal 2: Comparing Cohen2017 models



- Significant effect of flanking base pairs provides evidence that potential differences in optimized configuration of are in fact caused by dimer vs tetramer models
  - Not by other computational factors

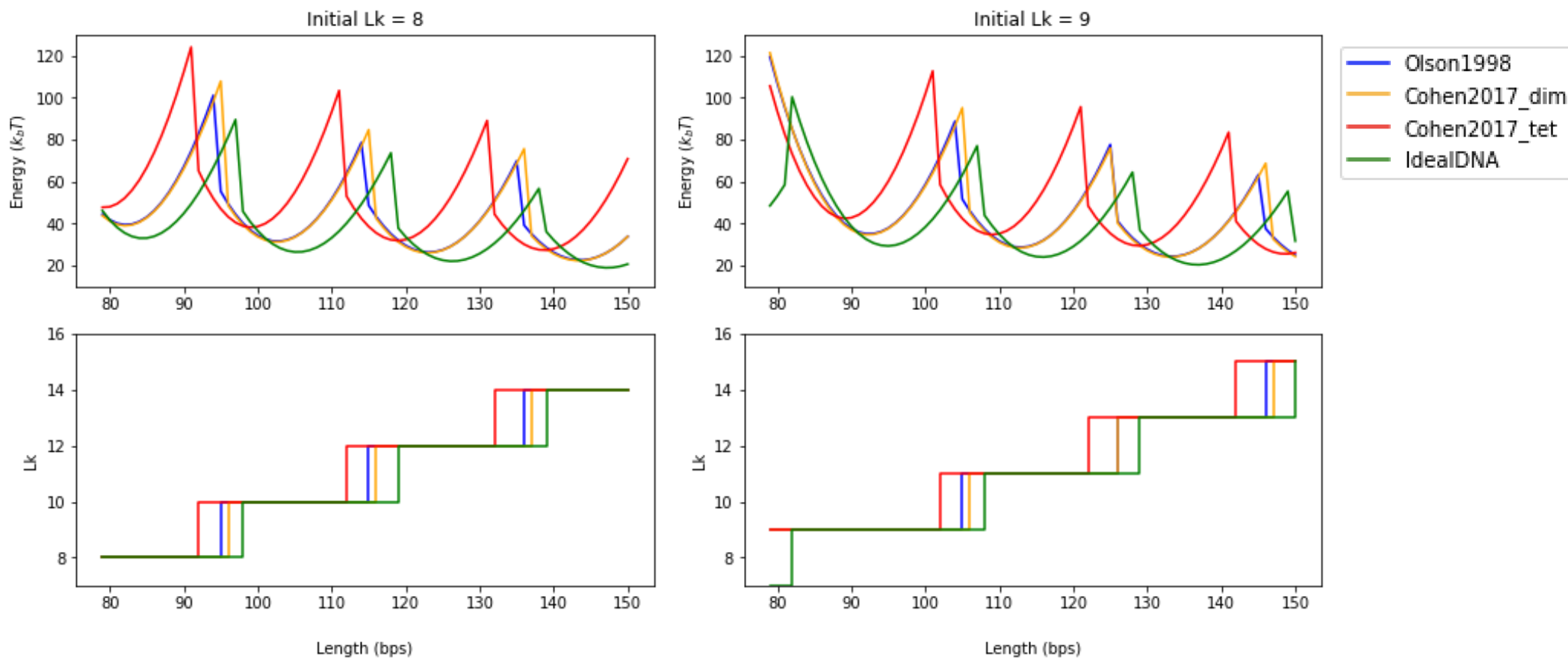
# Goal 3: Optimizing Minicircles

---

## Experimental Setup:

- Choose tetramer that is 1) easily repeatable and 2) intrinsic parameters vary significantly from intrinsic parameters of inner dimer
  - AAAA -  $1.5^\circ$  increase in intrinsic twist for AAAA tetramer compared to AA dimer in Cohen2017 models
- Optimize DNA Minicircles of lengths 80-150 base pairs and with linking number 8 and 9

# Final Energy and Final Linking Number



# Conclusions

Goal 1: Created new tool that expands the possibilities for exploring sequence dependence in DNA structure

Goal 2: Provided further evidence that base pairs which flank inner dimer impact intrinsic parameters

Goal 3: Results of optimization confirm that the scope of sequence dependent models has non-trivial effects on final conformation of DNA minicircles

# Next Steps

---

- Use higher resolution data to recompute Cohen2017 models
- Optimize DNA minicircles with more varied sequences
  - Maybe use repeating AG, AAGG, AAAGGG, etc.
- First and last base pair steps in a linear/open piece DNA sequence do not correspond to a specific tetramer
  - Process trimer data and include it in tetramer-dependent models



# References

---

1. Lu, 2003, '3DNA: A Software package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures', *Nucleic Acid Research*
2. Olson, W. K. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences*. Published. <https://doi.org/10.1073/pnas.95.19.11163>
3. Clauvelin, Olson, 2021, 'Synergy between Protein Positioning and DNA Elasticity: Energy Minimization of Protein Decorated Minicircles', *Journal of Physical Chemistry*
4. Irobalieva, R. N., Fogg, J. M., Catanese, D. J., Sutthibutpong, T., Chen, M., Barker, A. K., Ludtke, S. J., Harris, S. A., Schmid, M. F., Chiu, W., & Zechiedrich, L. (2015). Structural diversity of supercoiled DNA. *Nature Communications*, 6(1). <https://doi.org/10.1038/ncomms9440>
5. Pasi, M., Maddocks, J. H., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham, T., Dans, P. D., Jayaram, B., Lankas, F., Laughton, C., Mitchell, J., Osman, R., Orozco, M., Pérez, A., Petkevičiūtė, D., Spackova, N., Sponer, J., Zakrzewska, K., & Lavery, R. (2014).  $\mu$ ABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Research*, 42(19), 12272–12283. <https://doi.org/10.1093/nar/gku855>



# Acknowledgments

---

Thank you to the DIMACS REU 2021 program at Rutgers University, supported by NSF grant CCF-1852215. And thank you to Dr. Olson and Robert Young for their guidance!